Overview of the Agent-based Modeling and Behavior Representation (AMBR) Model Comparison Project

Kevin A. Gluck
Research Psychologist
Air Force Research Laboratory
6030 S. Kent St.
Mesa, AZ 85212
480-988-6561 x-234
kevin.gluck@williams.af.mil

Richard W. Pew
Principal Scientist
BBN Technologies
Cambridge, MA 02138
617-873-3557
pew@bbn.com

Keywords:
cognitive modeling, human behavior representation, multi-tasking, HLA

ABSTRACT: *The Air Force Research Laboratory (AFRL) is sponsoring a series of human performance model comparisons under the Agent-based Modeling and Behavior Representation (AMBR) Model Comparison Project. The first comparison challenged the modelers to build human cognitive models of multiple task management and attention sharing, embodied in the behavior of an air traffic controller operating in a simplified ATC task. The second comparison involved modifying the human behavior representation (HBR) models and the simulation environment to operate as an HLA Federation. BBN Technologies, as the model comparison moderator, designed the experiments, provided the simplified ATC simulation environment in which the models run, and collected data on actual human operators performing the task. CHI Systems, Soar Technology, Carnegie Mellon University, and a team from AFRL's Logistics and Sustainment Division built the HBR models. This paper is the first of several presented as part of the AMBR Symposium at the 10th Annual Computer-Generated Forces and Behavior Representation Conference, and it is intended as an introduction to that set. The paper provides an overview of AFRL's rationale for investing in this line of research, makes explicit the AMBR Model Comparison's goals, describes the iterative model comparison structure chosen for accomplishing those goals, and concludes with an introduction to the first two rounds of the project.*

1. Background and Rationale

In recent years, the Human Effectiveness directorate of the Air Force Research Laboratory (AFRL/HE) has increased its investment in science and technology for human behavior representation. This increase has occurred as a result of a convergence of evidence in the mid-to-late 1990's that there was an existing and future need for increased realism in models of human and organizational behavior for use in military simulations.

1.1 The National Research Council Report

At the request of the Defense Modeling and Simulation Office (DMSO), the National Research Council (NRC) established a panel to review the state of the art in human behavior representation as applied to military simulations. The panel consisted of leading experts in individual behavior, organizational behavior, decision making, human factors, computational modeling, and military simulations. Results of the panel's efforts were published in a 1998 book in which they concluded that "The modeling of cognition and action by individuals and groups is quite possibly the most difficult task humans have yet undertaken. Developments in this area are still in their infancy." [1, p. 341-342].

Having established the need for more research in this area, the panel suggested short-, intermediate-, and long-term goals for stimulating progress. These were intended to serve as suggested research directions for DMSO and other DoD agencies with an interest in modeling and simulation. A short-term suggestion was for increased human performance data collection. Real-world,

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
MAY 2001	**Conference Proceedings**	**01-01-2000 to 31-04-2001**

4. TITLE AND SUBTITLE
Overview of the Agent-Based Modeling and Behavior Representation (AMBR) Model Comparison Project

5a. CONTRACT NUMBER
F33615-99-C-6002

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63231F

6. AUTHOR(S)
Kevin Gluck; Richard Pew

5d. PROJECT NUMBER
4923

5e. TASK NUMBER
04

5f. WORK UNIT NUMBER
49230401

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RHA,Warfighter Readiness Research Division,6030 South Kent Street,Mesa,AZ,85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER
AFRL; AFRL/RHA

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/RHA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-AZ-PR-2001-0002

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
In Proceedings of the 10th Computer-Generated Forces/Behavioral Representation (CGF-BR) Conference, held 14-17 May 01 in Norfolk VA

14. ABSTRACT

The Air Force Research Laboratory (AFRL) is sponsoring a series of human performance model comparisons under the Agent-based Modeling and Behavior Representation (AMBR) Model Comparison Project. The first comparison challenged the modelers to build human cognitive models of multiple task management and attention sharing, embodied in the behavior of an air traffic controller operating in a simplified ATC task. The second comparison involved modifying the human behavior representation (HBR) models and the simulation environment to operate as an HLA Federation. BBN Technologies, as the model comparison moderator, designed the experiments, provided the simplified ATC simulation environment in which the models run, and collected data on actual human operators performing the task. CHI Systems, Soar Technology, Carnegie Mellon University, and a team from AFRL?s Logistics and Sustainment Division built the HBR models. This paper is the first of several presented as part of the AMBR Symposium at the 10th Annual Computer-Generated Forces and Behavior Representation Conference, and it is intended as an introduction to that set. The paper provides an overview of AFRL?s rationale for investing in this line of research, makes explicit the AMBR Model Comparison?s goals, describes the iterative model comparison structure chosen for accomplishing those goals, and concludes with an introduction to the first two rounds of the project.

15. SUBJECT TERMS

Human performance model; Agent-based modeling and behavior repreentation; AMBR; Model comparison; Human cognitive models; Task management; Attention sharing; Air Traffic Controller; Human behavior representation; HBR; Simulation environments; Human operators; Symposium;

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**15**	

wargame, and laboratory data all can and should be used to constrain the development of new and improved human behavior models. A second short-term suggestion was for the development of human model accreditation procedures, as none existed (and they still do not today). In the intermediate-term, the panel recommended that substantial resources be allocated for sustained model development in focused areas of interest to the DoD. In the long-term, the panel suggested support for theory development and basic research in areas such as decision making, situation awareness, learning, and organizational modeling.

1.2 The AMBR Project

Armed with these recommendations, AFRL/HE created a cognitive process modeling initiative for improving the realism of human behaviors as represented in military simulation environments, such as a Joint Synthetic Battlespace (JSB).

A JSB is a federation of virtual and constructive simulations and actual operational equipment networked together to perform an emulation of an operational mission. A JSB is used by warfighters to develop tactics and doctrine, formulate operational plans, assess war-fighting situations, and to train for war.

In order to improve the operational realism of the JSB and reduce the cost of conducting such simulations, AFRL/HE is using its new cognitive process modeling initiative, called the Agent-based Modeling and Behavior Representation (AMBR) Project, to develop and demonstrate new simulation technology in three areas: (1) the modeling of command and control (C2) echelons, (2) modeling the performance of technical controllers or support cell operations, and (3) the simulation of complex human behavior.

These three technology areas are all being pursued within the context of the AMBR Project, but AMBR is functionally divided into two research streams. One research stream focuses on developing and demonstrating technology at major Air Force exercises. This encompasses technology areas (1) and (2) above, and is not the topic of this paper.

The second research stream, which *is* the topic of this paper, focuses on the simulation of complex human behavior. We refer to this as the AMBR Model Comparison Project because multiple contractors receive funding simultaneously to model the same challenging cognitive phenomena, with the explicit intent of comparing and contrasting the design and predictive accuracy of the models at the end of each round of the project. An independent Moderator team from BBN Technologies has a contract with the Air Force to support the model comparisons.

2. AMBR Model Comparison Goals

There are three goals motivating the AMBR Model Comparison, all of which bear a striking resemblance to the recommendations made by the NRC panel mentioned earlier.

2.1 Goal 1: Advance the State of the Art

The first goal is to advance the state of the art in cognitive and behavioral modeling. This goal is consistent with the spirit of the entire set of recommendations from the NRC panel, since their recommendations were explicitly intended as a roadmap for improving human and organizational behavior modeling. With this goal in mind, we have devised a model comparison process (described in Section 3) that provides a motivation and opportunity for human modelers to extend and test their modeling architectures in new ways. As should be apparent in the modeling teams' papers for this Symposium, there is ample evidence just from Round 1 of the AMBR Model Comparison that these modeling architectures were being challenged, and in some cases improved, as a direct result of their participation in this project.

2.2 Goal 2: Develop Mission-Relevant HBR Models

The second goal is to develop HBR models that are mission-relevant, and therefore provide possible transition opportunities. This is consistent with the NRC panel recommendation to support model development in focused areas of interest to the DoD. An example transition possibility, higher fidelity human behavior models for the JSB, was mentioned earlier.

2.3 Goal 3: Make Tasks, Models, and Data Available

The third goal is to make all of the research tasks, human behavior models, and human process and outcome data available to the public. This is consistent with the NRC panel recommendation for increased collection and dissemination of human performance data. There is an AMBR website[1], but currently it is limited to a brief description of the project and downloadable files from the AMBR presentations at the Annual Meeting of the

[1] https://www.williams.af.mil/html/ambr.htm

Human Factors and Ergonomics Society. The intention going forward is to scale up the content of the website so the research task environment and models are available as well, either directly or via links.

3. AMBR Model Comparison Process

Round 1 of the AMBR Model Comparison involved the following steps:

(1) Identify the modeling goals for this iteration – what cognitive/behavioral capabilities should be stressed?

(2) Select a task domain that requires the capabilities identified in (1) and that is of relevance to AF modeling and simulation needs.

(3) Borrow/Modify/Create a simulation of the task domain which either a human-in-the-loop or a human performance model can operate.

(4) Moderator team collects and disseminates human performance data.

(5) Modeling teams develop models that attempt to replicate human performance when performing the task.

(6) Expert panel convenes to compare and contrast the models that were developed and the underlying architectures that support them.

(7) Share the results and lessons learned with the scientific community, to include making available the simulation of the task domain and the human performance data.

As of this writing, the 1st round is complete and the 2nd round is nearing completion. We describe these briefly below.

3.1 Round 1 (Multi-Tasking)

3.1.1 The Focus

The modeling focus for Round 1 was multiple task management, because this area represents a capability that is not widely available in existing models or modeling architectures, and because more knowledge regarding how to represent this capability provides an opportunity to improve the fidelity of future computer-generated forces (CGF's). It was up to the Moderator (BBN) to select a task for simulation that emphasized multiple task management.

3.1.2 The Task

Two approaches, representing ends of a continuum of intermediate possibilities, were considered. A task could be selected that was of operational interest, realistic complexity, and required highly trained operators to be our participants. Or the task could be highly abstracted, almost like a video game, that anyone could be expected to learn, but attempted to capture the task management requirements that were sought.

Clearly the first alternative would have greater practical significance and be more challenging from a modeling perspective. However, it would require extensive knowledge acquisition on the part of each development team, an investment that would detract from the time and effort that could be put into the modeling itself. The moderator could supply that knowledge, but it is well known that first hand knowledge is really required in order to address all the context-sensitive requirements. An overlay on this debate was whether the developers would be required to model experienced operators or novice operators. There were strong arguments against modeling novices, because the likely variability they would produce in the data would mask the behaviors we were trying to measure.

Using a task of realistic complexity also had implications for the Moderator team, which had limited resources for collecting data. Either they would have had to identify and recruit experienced operators from the domain under study, or invest in a very extensive period of training.

As a practical matter, the Moderator opted to use the highly-abstracted version of an air traffic control (ATC) task and utilized participants who had played a lot of video games, but had no previous experience with this task. Stable data were obtained from novice human participants in four-hour sessions and the modelers were able to develop the requisite knowledge based on their own experience or by testing a small set of previously untrained participants themselves. The paper by Tenney and Spector [2] contains considerable additional detail regarding the task, as well as human performance data, and the results of the comparison. The paper by Deutsch and Benyo [3] provides information about the underlying simulation environment (D-OMAR) used for both human-in-the-loop data collection and HBR model testing.

3.1.3 The Modeling Teams

Four modeling teams participated in Round 1. Two (CHI Systems and Soar Technology) were selected as part of the competitive bidding process at the beginning of

AMBR Round 1. A team from Carnegie Mellon University joined Round 1 in mid-course, with funding from the Office of Naval Research (ONR). Finally, a fourth modeling team, this one from the AFRL/HE Logistics and Sustainment Division, participated with no external funding. Each of these modeling teams has contributed a paper summarizing their accomplishments to date in the AMBR Model Comparison. We direct the reader to these papers (found in this volume) for information on the architectures they used, the models they developed, and their human performance predictions.

3.2 Round 2 (Icarus Federation)

In Round 2 of the AMBR Model Comparison Project, DMSO sponsored the conversion of the simulation environment and models from Round 1, so that they are compliant with the High-Level Architecture (HLA).

Goals for Round 2 include the following:

- Develop an HLA-compliant testbed for research in human behavior representation (HBR)
- Assess the adequacy of the HLA for supporting HBR research
- Assess the adequacy of DMSO's FEderation Development and Execution Process (FEDEP) as a framework for creating and running federations for HBR research

Since air traffic control is the simulation domain, the name of the testbed federation pays homage to a mythological analogy for air traffic control – Icarus (his father warned him not to fly too close to the sun). Figure 1 displays the federates of the Icarus Federation.

The Icarus Federation is the first documented application of the FEDEP to create a federation for HBR research. Thus, in AMBR Round 2 we are seeking to identify the compatibility (or incompatibility) of the techniques and tools developed to support federation development, execution, and management with models of perception, cognition, and motor movement. Tenney and Spector [2] provide data comparing human and model performance in the HLA and non-HLA versions of the simulation, while the Feinerman, Prochnow, and King [4] describe the process and lessons learned in transitioning the simulation and models from Round 1 to the Icarus Federation.

4. References

[1] Pew, R. W., & Mavor, A. S. (1998). *Modeling Human and Organizational Behavior: Application to Military Simulations*. Washington, D. C.: National Academy Press.

[2] Tenney, Y. J., & Spector, S. L. (2001). Comparisons of HBR Models with Human-in-the-loop Performance in a Simplified Air Traffic Control Simulation with and without HLA Protocols: Task Simulation, Human Data and Results. In *Proceedings of the 10th Annual CGF-BR Conference.*

[3] Deutsch, S., & Benyo, B. (2001). The D-OMAR Simulation Environment, including Non-HLA and HLA Federate Architectures. In *Proceedings of the 10th Annual CGF-BR Conference.*

[4] Feinerman, L. E., Prochnow, D. L., & King, R. A. (2001). Icarus, an HLA Federation of HBR Models. In *Proceedings of the 10th Annual CGF-BR Conference.*

Author Biographies

KEVIN GLUCK is a research psychologist at the Air Force Research Laboratory's Warfighter Training Research Division in Mesa, AZ. He is the program manager for AFRL's AMBR Model Comparison Project, and AFRL's POC for development and testing of the Icarus Federation. Dr. Gluck earned a B.A. in Psychology from Trinity University in 1993, an M.S. in Cognitive Psychology from Carnegie Mellon University in 1997, and a Ph.D. in Cognitive Psychology from Carnegie Mellon University in 1999.

RICHARD PEW is Principal Scientist at BBN Technologies LLC, a unit of the Verizon Technology Organization in Cambridge Massachusetts. Dr. Pew holds a bachelors degree in Electrical Engineering from Cornell University (1956), a master of arts degree in Psychology from Harvard University (1960) and a PhD in Psychology with a specialization in Engineering Psychology from The University of Michigan (1963). He has 35 years of experience in human factors, human performance and experimental psychology as they relate to systems design and development. Throughout his career he has been involved in the development and utilization of human performance models and in the conduct of experimental and field studies of human performance in applied settings. He spent 11 years on the faculty of the Psychology Department at Michigan where he was involved in human performance teaching, research and consulting before moving to BBN in 1974. His current research interested include the impact of automation on human performance, human-computer interaction and human performance modeling. He is the author or co-author of over 90 book chapters, research papers, conference proceedings, and technical reports.

Overview of the Agent-based Modeling and Behavior Representation (AMBR) Model Comparison Project

15 May 2001

Kevin A. Gluck
Air Force Research Laboratory

Richard W. Pew
BBN Technologies

Overview

- Project Origins
- Project Goals
- Round 1 (Multi-tasking)
- Round 2 (Icarus)

AMBR Origins

1996 – DMSO requests that the National Research Council (NRC) establish a panel to assess the state of the art in human and organizational behavior modeling

1998 – Panel conclusions and recommendations published as a book (Pew & Mavor, Eds.)

Primary Conclusion: "The modeling of cognition and action by individuals and groups is quite possibly the most difficult task humans have yet undertaken. Developments in this area are still in their infancy."

Some Recommendations:

More human performance data collection and dissemination

Substantial resources for sustained model development in focused areas of interest to the DoD

Support for theory development and basic research in areas such as decision making, situation awareness, learning, and organizational modeling

1999 - AFRL's Human Effectiveness directorate initiates the Agent-based Modeling and Behavior Representation (AMBR) Project

AMBR Organization

Two Research Tracks

PRACTICAL DEMONSTRATIONS

Application of a relatively mature technology to a specific problem.

-agent-based intelligent mission controller node (IMCN)

-Improving the behavior of autonomous models in command post exercises (CPX's)

EMERGENT DEVELOPMENT

Comparison of different modeling approaches for replicating human performance and learning data.

-Round 1 (multi-tasking)

-Round 2 (Icarus)

-Rounds 3 and 4 (category learning)

AMBR Model Comparison

AMBR Model Comparison Goals

- Advance the state of the art in cognitive and behavioral modeling
- Develop models that are relevant to identifiable Air Force modeling and simulation needs
- Create repository for simulation environments and human behavior data accessible to future modelers

Round 1: Multi-tasking

Modeling Focus

Multi-tasking (modeling human successes and shortcomings in coping with the demands of concurrent tasks)

Domain

Simplified version of enroute Air Traffic Control, in which the goal is to manage the central airspace

Moderator **BBN** (D. Pew, S. Deutsch, Y. Tenney, S. Spector, B. Benyo)

Participants

CHI Systems (COGNET/iGEN)

Wayne Zachary, Tom Santarelli, Joan Ryder, Jim Stokes, Dan Scolero

Soar Technologies (EPIC-Soar)

Ron Chong

Carnegie Mellon (ACT-R; funded by ONR)

Christian Lebiere, John Anderson, Dan Bothell

AFRL (D-COG; *pro bono*)

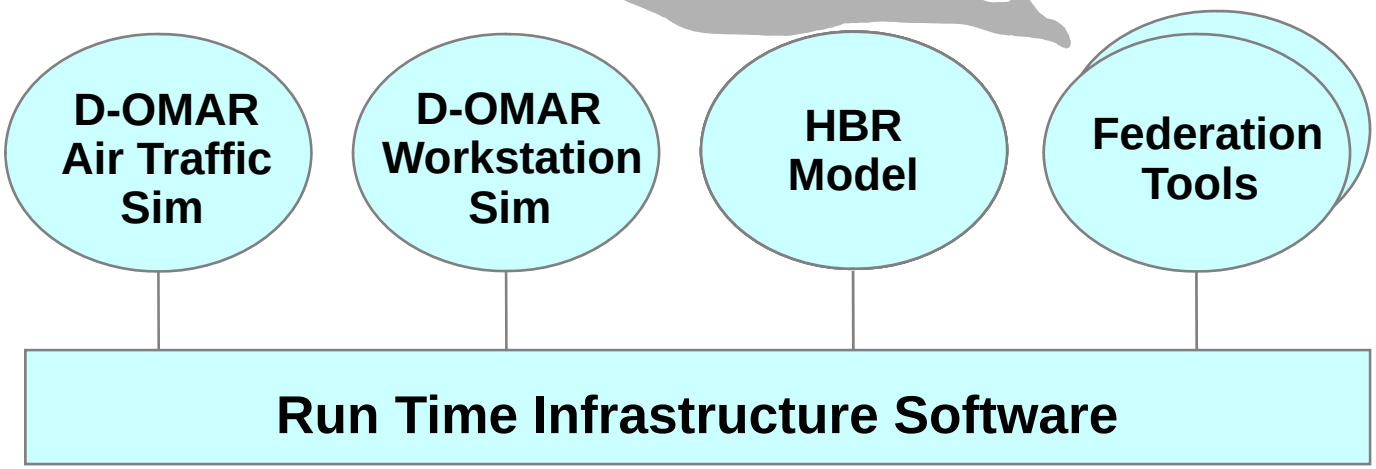
Bob Eggleston, Mike Young, Katherine McCreight

Architectures Differ ...

- in maturity
- in purpose
- in theoretical/empirical orientation

Round 2: Icarus Federation

DMSO-funded conversion of ATC task to HLA testbed federation for human behavior representation (HBR) studies



Icarus Federation

- **Same Domain**
- **Same Participants**

PLUS …

The MITRE crew

(Laura Feinerman, David Prochnow, Ron King)

Icarus Federation Research Goals

- **Create proof-of-principle testbed for Human Behavior Representation (HBR) research using the High-Level Architecture (HLA).**

- **Evaluate use of the HLA in human-in-the-loop (HITL) simulation and HBR model development.**

- **Evaluate generalizability of the Federation Development and Execution Process (FEDEP) in an HBR context.**

Now let's . . .

. . . get on with the show!!